CLAIMS

What is claimed is:

1. A method of using statistical analysis of genetic data to determine likely genetic regions for a recessive genetic disease or trait, comprising the steps of:

obtaining actual genotype data for one or more affected people with the genetic disease or trait in a population, for their parents, or for the affected people and their parents;

obtaining estimated genotype data for the population; and

analyzing the actual and estimated genotype data to find a region in genomes of the affected people that includes markers exhibiting particular homozygous pairs of alleles more frequently than would occur randomly, wherein the step of analyzing further comprises:

determining a set of scores under various assumptions for each marker in the genotype data relative to each person for which actual genotype data was determined;

merging the scores to arrive at a merged score for each marker; and

determining a region of markers that has a high run of merged scores.

2. A method as in claim 1, wherein the population is a relatively inbred population with a higher occurrence of the genetic disease or trait than a more general population.

3. A method as in claim 2, wherein the particular homozygous pairs of alleles are autozygous alleles descended from a founder of the genetic disease or trait in the relatively inbred population.

4. A method as in claim 3, wherein a score for a marker represents a comparison of a likelihood of observing the marker given that people with the genetic disease or trait are autozygous at the marker versus a likelihood of observing the marker given that alleles for the marker are independent of the genetic disease or trait.

5. A method as in claim 4, wherein a marker receives a higher score from one form of homozygosity versus another form of homozygosity, with the form receiving the higher score being more likely to be associated with the genetic disease or trait.

6. A method as in claim 5, wherein the merged scores are placed in an array ordered by a chromosomal order of markers associated with the scores.

7. A method as in claim 6, wherein the region of markers that has the high run of merged scores has the highest run of merged scores in the array; and

wherein the region of markers with the highest run of merged scores is found by determining a consecutive portion of the array that has the highest sum.

8. A method as in claim 6, wherein the region of markers that has the high run of merged scores is found by computing all sums of a predetermined fixed number of adjacent elements in the array and comparing the sums.

9. A method as in claim 6, further comprising the step of determining one or more additional regions of markers that have high runs of merged scores.

10. A method as in claim 9, further comprising the step of locating a statistically significant gap in the scores for non-overlapping regions, wherein regions having scores above the gap are determined to be the one or more additional regions of markers.

11. A method of analyzing actual and estimated genotype data, with the actual genotype data obtained for one or more affected people with the genetic disease or trait in a population, for their parents, or for the affected people and their parents, and with the estimated genotype data obtained for the population, the method performed to find a region in genomes of the affected people that includes markers exhibiting particular homozygous pairs of alleles more frequently than would occur randomly, the method comprising:

determining a set of scores under various assumptions for each marker in the genotype data relative to each person for which actual genotype data was determined;

merging the scores to arrive at a merged score for each marker; and

determining a region of markers that has a high run of merged scores.

12. A method as in claim 11, wherein the population is a relatively inbred population with a higher occurrence of the genetic disease or trait than a more general population.

13. A method as in claim 12, wherein the particular homozygous pairs of alleles are autozygous alleles descended from a founder of the genetic disease or trait in the relatively inbred population.

14. A method as in claim 13, wherein a score for a marker represents a comparison of a likelihood of observing the marker given that people with the genetic disease or trait are autozygous at the marker versus a likelihood of observing the marker given that alleles for the marker are independent of the genetic disease or trait.

15. A method as in claim 14, wherein a marker receives a higher score from one form of homozygosity versus another form of homozygosity, with the form receiving the higher score being more likely to be associated with the genetic disease or trait.

16. A method as in claim 15, wherein the merged scores are placed in an array ordered by a chromosomal order of markers associated with the scores.

17. A method as in claim 16, wherein the region of markers that has the high run of merged scores has the highest run of merged scores in the array; and

wherein the region of markers with the highest run of merged scores is found by determining a consecutive portion of the array that has the highest sum.

18. A method as in claim 16, wherein the region of markers that has the high run of merged scores is found by computing all sums of a predetermined fixed number of adjacent elements in the array and comparing the sums.

19. A method as in claim 16, further comprising the step of determining one or more additional regions of markers that have high runs of merged scores.

20. A method as in claim 19, further comprising the step of locating a statistically significant gap in the scores for non-overlapping regions, wherein regions having scores above the gap are determined to be the one or more additional regions of markers.

21. An apparatus including:

a processor;

input and output interfaces; and

a memory storing instructions executable by the processor to analyze actual and estimated genotype data, with the actual genotype data obtained for one or more affected people with the genetic disease or trait in a population, for their parents, or for the affected people and their parents, and with the estimated genotype data obtained for the population, the method performed to find a region in genomes of the affected people that includes markers exhibiting particular homozygous pairs of alleles more frequently than would occur randomly, the instructions including steps of: (a) determining a set of scores under various assumptions for each marker in the genotype data relative to each person for which actual genotype data was determined; (b) merging the scores to arrive at a merged score for each marker; and (c) determining a region of markers that has a high run of merged scores.